Analysis of optimal labelling problems and their application to image segmentation and binocular stereovision

Schlesinger M.I. [1], Flach B. [2]

[1] IRTC ITS, 40, prospect Academika Gluskova, 03680, Kyiv, Ukraine;
tel.: 266 62 08, E-mail: schles@image.kiev.ua

[2] Technische Universität Dresden, Fakultät Informatik, Institut für künstliche Intelligenz,
D-01062 Dresden, Deutschland
Email: bflach@ics.inf.tu-dresden.de

ABSTRACT

The approach to optimal labelling is described. Algorithms developed within the proposed approach are of polynomial complexity and defined on the whole class of labelling problems. Some labelling problems are processed so that no labelling is given and such result shall be interpreted as an answer "I do not know". However, if the algorithm produces a labelling it can be only optimal. Such feature of proposed algorithms distinguishes them essentially from known algorithms, which are either defined for some subclass of labelling problems or fulfil local improvements of labelling and do not provide the globally optimal solution.

1. FORMULATION OF THE OPTIMAL LABELLING PROBLEM

Let T be a finite set of pixels, $\Im \subset T \times T$ be a subset of pixel pairs referred to as neighbours, K be a finite set of labels. A function of the form $\bar{k}: T \to K$ will be called a labelling, the set of all possible labellings will be denoted K^T. The label of the pixel t will be denoted $k(t)$.

Let for every pair $tt' \in \Im$ of neighbours a function $g_{tt'}: K \times K \to R$ be defined as well as a function $q_t: K \to R$ for every pixel $t \in T$. The quality of the labelling $\bar{k} \in K^T$ is defined as

$$G(\bar{k}) = \sum_{tt' \in \Im} g_{tt'}\left(k(t), k(t')\right) + \sum_{t \in T} q_t\left(k(t)\right) \qquad (1)$$

The optimal labelling problem consists in constructing an algorithm which gets input data

$$z = \left\langle T, \Im, K, (g_{tt'} \mid tt' \in \Im), (q_t \mid t \in T) \right\rangle$$

and finds the best labelling

$$\bar{k}^* = \arg\max_{\bar{k} \in K^T} \left[\sum_{tt' \in \Im} g_{tt'}\left(k(t), k(t')\right) + \sum_{t \in T} q_t\left(k(t)\right) \right]. \qquad (2)$$

Hereinafter we will sometimes omit the second sum in the expression (2), assuming that all numbers $q_t(k)$, $t \in T$, $k \in K$, are zeros. This does not constrict the class of problems under consideration if each pixel has at least one neighbour. Indeed, for each labelling $\bar{k} \in K^T$ the equality

$$\sum_{tt' \in \Im} g_{tt'}\left(k(t), k(t')\right) + \sum_{t \in T} q_t\left(k(t)\right) = G(\bar{k}) = \sum_{tt' \in \Im} g^*_{tt'}\left(k(t), k(t')\right) \qquad (3)$$

is valid, where

$$g^*_{tt'}(k, k') = g_{tt'}(k, k') + \frac{q_t(k)}{|N(t)|} + \frac{q_{t'}(k')}{|N(t')|} \qquad (4)$$

and $N(t)$ is the set of neighbours of the pixel t.

The set of problems of the form (2) is *NP*-complete. The known approaches to the problem can be divided into two groups. In the works of the first group some polynomially solvable subclass of labelling problems is specified with constrains either on the functions $g_{tt'}: K \times K \to R$ [1, 2, 3, 6] or the neighbourhood $\Im$ [5, 7]. Algorithms of the second group fulfil local step-by-step improvements of current labelling. Such algorithms are defined on the whole set of labelling problems, but some problems are not solved correctly: the algorithm may specify a locally unimprovable labelling which is not optimal. We propose another approach that distinguishes from the well-known ones as it was quoted in the Abstract.

2. DESCRIPTION OF THE APPROACH

2.1. Definition of trivial problems

Let $z = \left\langle T, \Im, K, (g_{tt'} \mid tt' \in \Im) \right\rangle$ be input data for a labelling problem

© UkrObraz'2002

$$\bar{k}^* = \operatorname*{argmax}_{\bar{k}\in K^T} \sum_{tt'\in\Im} g_{tt'}(k(t),k(t')).$$

Let the functions $\bar{g}_{tt'} : K\times K \to \{0,1\}$ be such that

$$\bar{g}_{tt'}(k,k') = 1 \ ,$$

$$\text{if } g_{tt'}(k,k') = \max_{k,k'} g_{tt'}(k,k') \ ,$$

$$\bar{g}_{tt'}(k,k') = 0 \ ,$$

$$\text{if } g_{tt'}(k,k') < \max_{k,k'} g_{tt'}(k,k') \ .$$

The problem $z = \langle T,\Im,K,(g_{tt'} \mid tt'\in\Im)\rangle$ is called trivial if there exists a labelling $\bar{k}^{\circ}$ such that

$$\mathop{\&}_{tt'\in\Im} \bar{g}_{tt'}(k^{\circ}(t),k^{\circ}(t')) = 1 .$$

Obviously, the labelling k° is optimal in this case. Indeed, for each labelling $\bar{k}' \in K^T$ the following chain of equalities and inequality

$$\begin{aligned} G(\bar{k}') = \sum_{tt'\in\Im} g_{tt'}(k'(t),k'(t')) \le \\ \le \sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k') = \\ = \sum_{tt'\in\Im} g_{tt'}(k^{\circ}(t),k^{\circ}(t')) = G(\bar{k}^{\circ}) \end{aligned} \qquad (5)$$

is valid. The number $\sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k')$ in the chain does not depend on the labelling, it depends only on the problem z. This characteristic of the problem will be called the problem potential and denoted $\Phi(z)$,

$$\Phi(z) = \sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k') . \qquad (6)$$

2.2. Equivalent problems

Two problems

$$z_1 = \langle T,\Im,K,(g^1_{tt'} \mid tt'\in\Im)\rangle$$

and

$$z_2 = \langle T,\Im,K,(g^2_{tt'} \mid tt'\in\Im)\rangle$$

are called equivalent if for each labelling $\bar{k} \in K^T$ the equality

$$\sum_{tt'\in\Im} g^1_{tt'}(k(t),k(t')) = \sum_{tt'\in\Im} g^2_{tt'}(k(t),k(t')) \qquad (7)$$

is valid. This definition is not constructive because recognition of equivalency requires checking the $|K|^{|T|}$ equalities (7). The following theorem defines the equivalency in the constructive way.

Theorem 1. If the neighbourhood $\Im$ forms a connected graph, then the problems

$$\langle T,\Im,K,(g^1_{tt'} \mid tt'\in\Im)\rangle \ ,$$
$$\langle T,\Im,K,(g^2_{tt'} \mid tt'\in\Im)\rangle$$

are equivalent if and only if there exists an array of numbers $\varphi_{tt'}(k),\ t\in T,\ t'\in N(t),\ k\in K$, which satisfy the equalities

$$\left.\begin{aligned} \varphi_{tt'}(k) + \varphi_{t't}(k') = g^2_{tt'}(k,k') - g^1_{tt'}(k,k'), \\ t\in T,\ t'\in N(t),\ k\in K,\ k'\in K; \\ \sum_{t'\in N(t)} \varphi_{tt'}(k) = 0,\ t\in T,\ k\in K. \end{aligned}\right\} \qquad (8)$$

■

2.3. Transformation of the non-trivial problem into trivial

The proposed approach consists in searching for a trivial equivalent for the problem under solution. Certainly, it can be done only if such a trivial equivalent exists. Implementation of this idea is based on the following considerations.

Let Z be an equivalency class, z being a problem of this class; let $\bar{k}$ be some labelling. The problem potential $\Phi(z)$ does not depend on the labelling and the labelling quality $G(\bar{k})$ does not depend on the problem of the class Z. The inequality

$$\Phi(z) \ge G(\bar{k})$$

is evident (see (3)). If z^* is a trivial problem and $\bar{k}^*$ is an optimal labelling then the inequality

$$\Phi(z^*) = G(\bar{k}^*)$$

is evident too (see (3)). It means that the following theorem is valid.

Theorem 2. If z^* is a trivial problem then

$$z^* = \arg\min_{z \in Z} \Phi(z),$$

where Z is a set of problems which are equivalent to z^*. ■

The following theorem, which is inverse in certain sense, is valid too.

Theorem 3. If a class of equivalent problems includes at least one trivial problem then any problem

$$z = \arg\min_{z \in Z} \Phi(z)$$

is trivial. ■

The following theorem shows the constructive way for searching for a problem with the minimal potential.

Theorem 4. For each problem z' there exists a problem

$$z'' = \arg\min_{z \in Z} \Phi(z),$$

where Z is the class of problems, equivalent to z'. This problem is a solution of the following linear programming problem:

minimise $$\sum_{tt' \in \Im} h(tt') \quad (9)$$

under conditions

$$\left.\begin{aligned} & h(tt') \geq g''_{tt'}(k,k'),\ tt' \in \Im,\ k \in K,\ k' \in K, \\ & \varphi_{tt'}(k) + \varphi_{t't}(k') = g''_{tt'}(k,k') - g'_{tt'}(k,k'), \\ & \qquad t \in T,\ t' \in N(t),\ k \in K,\ k' \in K, \\ & \sum_{t' \in N(t)} \varphi_{tt'}(k) = 0,\ t \in T,\ k \in K. \end{aligned}\right\} \quad (10)$$

■

The fulfilled analysis allows to formulate the following approach to optimal labelling searching.

1. Find a problem equivalent to the initial one, which minimises the problem potential.
2. Check whether the found problem is trivial or not.
3. If YES, declare any solution to the trivial problem to be a solution to the initial problem.
4. If NO, choose no solution and interpret such an outcome as "I DO NOT KNOW" answer.

We will show how some image segmentation and binocular stereovision problems are reduced to optimal labelling searching. Moreover, owing to the peculiarities of these problems the do-not-know answer is impossible. It means that problems of such class admit an exact solution.

3. IMAGE SEGMENTATION

Let T be a set of pixels, $\Im$ be a neighbourhood, X be a set of signal values observed in each pixel, K be a set of segment names. An image is a function $\bar{x}: T \to K$ and a segmentation is a function $\bar{k}: T \to K$.

A priori quality is defined for each segmentation; it is based on the intuitive idea that two neighbouring pixels most likely belong to the same segment. This can be expressed so that the a priori quality of the segmentation $\bar{k}$ is the sum

$$\sum_{tt' \in \Im} g_{tt'}\big(k(t), k(t')\big),$$

where $$\left.\begin{aligned} & g_{tt'}(k,k') = \alpha > 0, \ \text{if } k = k', \\ & g_{tt'}(k,k') = 0, \qquad \text{if } k \neq k'. \end{aligned}\right\} \quad (11)$$

For each pair $\bar{x}, \bar{k}$ (image - segmentation) a similarity measure is defined

$$\sum_{t \in T} q_t\big(k(t), \bar{x}\big). \quad (12)$$

The numbers $q_t(k, \bar{x})$, $k \in K$, $\bar{x} \in X^T$, in the sum (12) express to what extent a decision for the segment k in the pixel t goes with the image $\bar{x}$ under observation. We do not here go into problems on how these numbers must be reasonably chosen, since a possibility of constructive problem solution is not determined by the form of the functions q_t but the form of the local qualities $g_{tt'}$ given by the expression (11).

The segmentation problem is formulated as searching for a function $\bar{k}: T \to K$, which maximises the sum of its a priori quality and its similarity to the image $\bar{x}$.

4. BINOCULAR STEREOVISION

Let T be a finite set of points on a 2-D-plane. A surface is a function $\bar{k}: T \to K$, where $k(t)$ is a height of the surface above the point $t \in T$. Let a subset of admissible surfaces is chosen in the following way: for each pair $tt' \in \Im$ of neighbouring points a number $\Delta_{tt'}$ is specified, and a surface $\bar{k}$ is regarded as admissible if heights in neighbouring points differ by no more than $\Delta_{tt'}$. It means that admissibility of the surface $\bar{k}$ is defined by the sum

$$\sum_{tt' \in \Im} g_{tt'}\big(k(t), k(t')\big),$$

where $$g_{tt'}(k,k') = 0, \quad \text{if } |k-k'| \le \Delta_{tt'}, \\ = -\infty, \quad \text{if } |k-k'| > \Delta_{tt'}. \tag{13}$$

A surface is admissible if the sum (13) is equal to zero and inadmissible otherwise.

The surface $\bar{k}: T \to K$ is not directly observable. Instead, there are two images $\bar{x}_1$ and $\bar{x}_2$ under observation, which form a stereopair. On the basis of these images the numbers $q_t(k,\bar{x}_1,\bar{x}_2)$ are calculated, which specify, how well a decision that the height of the surface above the point t is k goes with the observable images $\bar{x}_1$ and $\bar{x}_2$. Degree of conformity of the surface $\bar{k}$ with the stereopair $\bar{x}_1$ and $\bar{x}_2$ is defined as the sum

$$\sum_{t\in T} q_t(k(t),\bar{x}_1,\bar{x}_2).$$

We do not here consider how the numbers $q_t(k,\bar{x}_1,\bar{x}_2)$ must be calculated, because it will be shown below that a possibility of the constructive problem solution is determined only by the form (13) of the functions $g_{tt'}$. The binocular stereovision problem consists in searching for such an admissible surface which shows the best conformity with the observations $\bar{x}_1$ and $\bar{x}_2$.

5. MONOTONOUS LABELLING PROBLEMS

The labelling problem is called monotonous if the set K is ordered and the numbers $g_{tt'}(k,k')$ satisfy inequalities

$$g_{tt'}(k_1,k_2') + g_{tt'}(k_2,k_1') \le g_{tt'}(k_1,k_1') + g_{tt'}(k_2,k_2') \tag{14}$$

for any $tt' \in \mathfrak{I}$ and $k_1 < k_2$, $k_1' < k_2'$.

Image segmentation into two segments and binocular stereovision problem in the above stated formulations are monotonous labelling problems.

It is worth mentioning that in monotonous problems only the form of functions $g_{tt'}$ is restricted and by no means the form of functions q_t. It was noted above that the numbers $q_t(k)$ can be set to zeros by means of changing the numbers $g_{tt'}(k,k')$ without loss of generality (see (4)). It is essential that after changing the numbers $g_{tt'}(k,k')$ by the formula (4) the problem remains monotonous. Moreover, any equivalent transformation of the problem preserves monotonicity in the sense of definition (7).

A remarkable attribute of monotonous problems consists in validity of the following theorem.

Theorem 5. For any monotonous problem there exists an equivalent trivial problem. ■

CONCLUSION

The approach to construction of optimal labelling algorithms is described, which are not defined on a subclass of labelling problems, but on the class of all possible problems, which is well known to be *NP*-complete. After processing some input problems such an algorithm may not output a labelling but a special answer, which must be understood as a denial of solution of exactly this problem, as an answer "*I do not know*". Essential advantage of such algorithms is, however, that the situation is excluded when an algorithm outputs a non-optimal labelling. Construction of such optimal labelling algorithms is feasible, this being the main scientific result of the given research. Besides the main result it is essential that the certain problem subclass is described, for which the answer "*I do not know*" cannot be given. Some image segmentation and binocular stereovision problems are included in this class, and their exact solutions can be thus obtained.

REFERENCES

[1] Boykov, Yu., Veksler, O., and Zabih, R., Fast Approximate Energy Minimization via Graph Cuts, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 11, November 2001.

[2] Ishikawa H., Geiger D., Segmentation by Grouping Junctions, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998.

[3] Kolmogorov, V., and Zabih, R., What Energy Functions Can Be Minimized via Graph Cuts?, in:A. Heyden et al. (ed.), Computer Vision - ECCV2002, LNCS 2352, Copenhagen, Denmark, May 2002.

[4] Schlesinger M.I. Sintaksicheskij analiz dvumernyh zritel'nyh signalov v uslovijah pomeh, in Russian, (Syntactic analysis of noisy images). Kybernetika, 4, 1976.

[5] Schlesinger, M.I., Matematiceskie sredstva obrabotki izobrazenij, in Russian (Mathematical tools for image processing). Naukova Dumka, Kiev, 1989.

[6] Schlesinger, M.I., and Flach, B., Some solvable subclasses of structural recognition problems, in: T. Svoboda (ed.), Proceedings of the Czech Pattern Recognition Workshop 2000, Praga 2000.

[7] Schlesinger, M.I., and Hlavac, V., Ten Lectures on Statistical and Structural Pattern Recognition, Computational Imaging and Vision, Kluwer Academic Publishers, Dordrecht / Boston / London, 2002.